Contribution of proteomics to colorectal cancer diagnosis

Marie-Alice Meuwis[1], Edouard Louis[2] and Marie-Paule Merville[1]

[1]*Clinical chemistry, CBIG, GIGA, CHU, University of Liège*

[2]*Hepatho Gastroenterology, CHU, University of Liège*

E-Mail: mpmerville@ulg.ac.be

Address:

Clinical chemistry

CHU, B23, +3

Sart Tilman

4000 Liège

Belgium

Key words: proteomic, colorectal cancer, serum profiling, biopsy, diagnosis

Abstract

Colorectal cancer is the second cause of cancer related death in developed countries. It is of major concern for public health authorities which are interested in large population screening for CRC, as early identification leads to decreased mortality. Unfortunately classical clinical diagnosis methods are too expensive to be applied for large screening. The development of rapid, low cost, easy and accurate tools for CRC diagnosis is needed. Biomarkers as sensitive than specific, able to achieve prognosis of CRC and correlated to clinical features still remain elusive. But the growing fields of new techniques like proteomic ones might be a suitable option to discover new specific biomarkers for CRC and for developing new tools of diagnosis. These techniques have already given new insights in various cancers and other diseases. This paper reviews the current knowledge in the fields of proteomics dedicated to CRC and more precisely to CRC diagnosis.

Résumé

Le cancer du colon est le deuxième cancer le plus mortel dans nos pays développés. Les autorités responsables de la santé publique souhaiteraient pouvoir effectuer des dépistages systématiques du cancer du colon, puisque un diagnostic précoce réduit les risques de mortalité. Malheureusement le coût des méthodes classiques de diagnostic est actuellement trop important pour envisager un dépistage systématique à grande échelle. Le développement de nouveaux outils diagnostiques, rapides, faciles, peu coûteux et efficaces pour le cancer du colon est donc nécessaire. Néanmoins, des biomarqueurs aussi sensibles que spécifiques, capables de diagnostiquer les stades précoces de cette maladie et corrélés aux méthodes cliniques classiques de diagnostic reste un idéal non atteint. Des techniques émergeantes et puissantes de protéomique peuvent constituer une option séduisante et adaptée afin de découvrir de nouveaux biomarqueurs spécifiques du cancer du colon et permettre le développement de nouveaux outils diagnostiques. Le nouveau champ de recherche qu'est la protéomique clinique et les technologies qu'elle exploite ont déjà porté leurs fruits dans le cas d'autres cancers et autres pathologies. Le but de ce papier est de rassembler et commenter les découvertes et progrès constants faits dans le domaine de la protéomique appliquée au cancer du colon et plus précisément au diagnostic de ce type de cancer.

Introduction

Despite the increasing knowledge on colorectal cancer etiology and the growing development of specific therapies and diagnosis tools, CRC is still the second cause of cancer related death in western countries. The early diagnosis of CRC or any preceding stages presenting colon abnormalities (including adenomatous polyps or plane lesions) is required to increase chances of survival [1]. Nowadays, CRC diagnosis still relies on classical clinical methods like colonoscopy, double contrast barium enema, or the recent Virtual Colonoscopy (or Computed Tomographic Colonography). This last one presents the advantages of being less invasive than endoscopic colonoscopy, necessitating less radiation than contrast barium enema and following some technical improvement might even be performed without any bowel preparation, tagging the stools with contrast agent. Nevertheless it sensitivity seems to be variable depending on case reports [2]. But, such techniques are to some extent risky (accidental perforations have been recognized both with endoscopic or virtual colonoscopy) and are too expensive to be used as large screening tools. Only a few molecular based diagnosis tests, performed in particular conditions, are recommended for CRC diagnosis and follow up: CEA (carcinoembryonic antigen) in blood or faeces, CA19-9 (gastrointestinal carbohydrate antigen 19-9), faecal occult blood test (FOBTs). As CRC early diagnosis is an important public health consideration, many governments (United States, Denmark, UK and Australia) are involved in studying, prospectively, on large cohorts of people the effect of the use of molecular diagnosis testing on CRC mortality. FOBTs, CEA and CA19-9 appear to be interesting in many CRC stages or are prognosis factors. CRC screening using FOBTs has been shown to decrease the CRC-related mortality in several large population studies [1]. However this technique is flawed by a significant proportion of false negative. CEA and CA19-19 have no demonstrated value as a screening or diagnostic tool. They are essentially usefull for purposes like monitoring recurrence after surgery and other treatments [3, 4]. But, their specificities and sensitivities taken solely are not satisfying. That is the reason why, beside these markers of tumours burden, numerous teams are assessing with classical strategies new molecules and their producing and downstream metabolic events, in order to implement the diagnosis of CRC [4].

Among these markers still under study, one might notice genetics factors like Ras family gene mutations, P53 mutations, microsatellite instability and chromosomal

translocations; molecules involved in angiogenesis, like VEGF, cell adhesion molecule; inflammatory related molecules, like Tissue Factor, S100A4, CRP and various components involved in many pathways, like nuclear matrix protein (NMP), Thymidylate synthase (TS) mRNA, some polyamins and the von Willebrand factor [4-9] . Unfortunately, these markers are not highly specific to CRC, as they are also found in other cancer types. Then, many authors propose to use several markers combined in multivariate model in order to obtain more specific and accurate diagnosis of CRC [10, 11].

In this context, the evolving techniques of proteomics are interesting to discover new biomarkers in various contexts. Clinical proteomics is an emerging area already reporting many advances in the field of diagnosis of various pathologies as cancer [12-14]. Hence, it allows the evaluation of particular proteomes through very sensitive technical approaches and arises with the possibility of monitoring and combining many biomarkers in one single test. Moreover, it shows high throughput screening capabilities. Among the vast ongoing publications reported to date, this paper will focus on proteomics contribution to the discovery of news potential specific CRC biomarkers and also on the development of new diagnosis tests for CRC, based on protein profiling.

1. General considerations about proteomic techniques

Several technologies like 2D-gel electrophoresis or others specialized in proteins profiling exist and are dedicated to the study of the comparison of proteins expressed in particular conditions by a tissue or present in a given body fluid. These allow the determination through accurate statistical methods, of some proteins or peptides differentially found present in one or the other compared sample. Two rational strategies exist in the current works reported. Generally, scientists study biomarkers of interest individually in details, purifying and identifying them, correlating either their proteomic observations with classical methods. The second strategy uses all the data collected in the profiles and designs many models of classification with robust bioinformatics and statistical tools. The protein signature obtained may therefore orientate the classification of the sample in a given category, without knowing the identity of all protein factors contributing to the final diagnosis. Nevertheless, due to the difficulty of replication of such integrative diagnostic or predictive profiles, the scientific community stay very cautious with this second option and prefers to establish

models of classifications with identified biomarkers, which are consistent with the etiopathology.

Samples of different origins can be compared ranging from model cell lines, body fluids (sera, plasma, urine, ascites, tears...) or cells derived from animal models and from patient tissues (biopsies, blood cells…). These samples are diluted or can even be fractionated in order to detect minor proteins. Moreover, by comparing patients subjected to various diseases managements, diets, over long periods of time, presenting different stages or grades of disease…, these proteomic techniques may provide valuable and complete data.

Clinical proteomic advances made and new insights brought by these techniques are reviewed in [15]-[16]. Techniques as various as 2D gel electrophoresis, profiling on SELDI-TOF-MS or MALDI-TOF-MS, MALDI imaging and combinations of several technologies like LC/MS-MS, 2DLC/MS-MS, ICAT, SILAC, nanotechnologies provides strong platforms with various capabilities. One advantage is in the power of mass spectrometry which have been developed and used primarily for assessing quality controls for example, in pharmaceutical industry. Mass spectrometry usually requires very small sample quantity. In addition to its sensitivity, it brings the information on mass very precisely, which is an element not available by classical techniques. This way giving access to possible posttranslational modifications or variant isotypes, differing in mass and which can be missed by other techniques like classical Western blotting or ELISA. Another strength of the overall strategy is the fractionation of samples. This preparation step is based on classical and well known methods of biochemistry and may be completely standardized and transposed to nanoscale, reducing time, variation and consumption of precious materials. Hence, proteomic evolving very rapidly towards more automation and reproducibility, it should soon satisfy robustness required in clinical diagnosis, predicting in the near future new developments in this area and new diagnosis solutions. The study by proteomic of pathologies as frequent as CRC is more and more popular judging by the numerous publications reported to date.

2. Proteomic on CRC

2.1. Serum profiling

The recently developed SELDI-TOF-MS technology is based on the binding of proteins on chemically activated groups coated on chip arrays. These on chip purified proteins are resolved with a Surface Laser Enhanced Desorption Ionisation Time Of Flight Mass

Spectrometer, which reconstitute profiles according to the panel of proteins detected [17, 18]. This technology has been used in several studies with serum of CRC patients with good success [19-22]. Authors use serum of patients suffering from CRC at different grade or stage (Dukes grades and stages I to VI). They prepared samples by running a pre fractionation directly on chip arrays and obtained different protein patterns as they utilised different nature of chips. They all used sophisticated bioinformatic machine learning algorithm and artificial neural networks classification tools to manage data profiles and propose models of classification based on a few selected potential biomarqueurs. These were able in these first studies to differentiate CRC from healthy patients and even differentiate grades or stages of CRC, with excellent sensitivity and specificity. Nevertheless, none of these teams have performed or yet published validation or blind test to confirm these results. More completed works have been performed with SELDI-TOF-MS on biopsies and were confirmed in sera [19]. Finally, another group interested in prostate cancer has reported a complete study, using SELDI-TOF-MS serum profiling and achieve accurate discrimination of patients. They validated successfully their models on several SELDI platforms, located in USA and Europe. They also advised modalities to control reproducibility with this technique, attesting of efforts made to answer general criticisms on SELDI-TOF-MS [20].

2.2. Patient's biopsy

Proteomic study of biopsy is probably the more documented one and is by far less criticised. Indeed, every patient may possibly be its own negative control, in the comparison of tumour tissue and not affected neighbouring one. But, the sensitive question relies more on the possibility to use lazer microdissected cells, separating tumour cells from contact epithelial cells or else the entire biopsy with every tissue types included [16].

Studies with the quite sensitive 2D-DIGE gel electrophoresis and identification of differentially expressed proteins by MALDI-TOF-MS have been recently reported [21], [22]. These works focus either on biopsies from tumour cells and adjacent non tumour tissue from CRC patients or on biopsies from patients suffering from CRC at different sites. They identified more than 32 proteins differentially present among groups, but only the last authors confirmed part of their results by specific immunoblotting or immunohistochemistry. Among other, they confirmed several variant isoforms and a phosphorylated form of vimentin, cytokeratin 8 and 19, calreticulin, apolipoprotein A1… This work revealed category and family of proteins found previously involved in carcinogenesis by classical genomics and in

other proteomic studies: MAT-1 (a protein identified from metastatic mammary adenocarcinoma cell line), cathepsin D and calreticulin [22]. However, such time consuming studies can only be run on a few samples at a time and the relevance of candidate biomarkers should be addressed in larger cohort of patients, to reach statistical significance.

In contrast to 2D-DIGE, SEDLI-TOF analysis can be performed on many sample lysates of tumour and non tumour tissue to obtain statistical significance. Many teams used this strategy with success allowing the discovery, in a rather small study, of a group of not yet identified peptides in the mass range of 3.4 to 3.6 kDa, which were increased in CRC tumour cells[23]. Moreover, two very elegant and complementary works have been performed by independent groups[24],[19]. They both reported in tumour cells and in sera, that CRC patients showed increased levels of α-defensin 1, 2, 3 peptides or HNP1-3, at mass 3372, 3442 and 3486Da. These observations were confirmed through other techniques like immunohistochemistry which gave the repartition of HNP1-3 in tumour cells and in normal surrounding tissues, as macrophages and fibroblasts of the mucus membrane indicating some potential roles of these peptides in the aetiology of many cancers and in inflammatory pathologies like inflammatory bowel disease. Of course, similar strategies can be transposed to study drugs metabolism and further follow up and management of CRC patients, like proposed in the evaluation of cytochrome P450 by proteomic approach [25].

Some authors reported the discovery of two proteins specifics for CRC development, using tumor microdessected cells from adenoma and normal colon cells. They identified HSP10 by proteomic profiling. This heat shock protein 10 was found increased in CRC adenoma and confirmed by direct immunolabelling of tissue sections and by Western blotting analysis on cell lysates [26]. They else reported the comparison of protein profiles generated on SELDI-TOF-MS with normal, adenoma and carcinoma microdissected cells of colon. Among proteins found differentially present, there was Calgizzarin (or S100A11), which could be further confirmed by specific immunodepletion from the original sample at expected peak position in spectra and by immunohistological analysis.

Recent developments of proteins microarray analysis have been realized, combining data from genomic and proteomic discovery. For example, in CRC diagnosis, Belluco and Lise tested the use of a specific phosphoprotein chip array. This includes the use of 29 specific antibodies targeting phosphoproteins known to be part of end point signalling cascades involved in carcinogenesis, like clived Caspase 3, Erb2, ERK or NF-kB P65. They utilized neoplastic microdissected cells from primary CRC, with and without liver metastases and liver metastases cells, to characterise the answers, on chip of these non invasive or

invasive stages of CRC. They used sophisticated bioinformatics and mathematical tools to associate different metabolic phosphoproteomic profiles with CRC stages [27]. They assume that microenvironment involves different signalling pathways that can be sensed with this type of techniques. Thus, similar tools might be developed assessing the biomarkers or signature of patients utilising for example the well characterised biomarkers proposed in proteomic study.

3. Emerging techniques and strategies

The field of proteomics combines technologies which are in constant evolution and improvement. For example, mass spectrometry instruments constructors which develop engines which are robust and efficient, automatized and easy to use. Some of these can be dedicated to various application like identification of protein, protein profiling, quantitative proteomics using methods like isotope coded affinity tag (ICAT), isotope tag for relative and absolute quantitation (iTRAQ™) or stable isotope labelling of amino acid in cell culture (SILAC) and application for imaging (MALDI imaging), which is also commercialised. Moreover, they may be combined with various scale purification platforms, microflow or nanoflow HPLC, enabling better separation of products for analysis of pure species. The nano technologies are of particular interest knowing recent developments in biomaterials designed for purification, a very important step for correct protein identification by sequencing. In addition, efforts have also been made to improve 2D gel based analysis with the 2D-DIGE technology, which uses more sensitive dyes than conventional silver staining or coomassie blue procedures. Indeed, different samples to be compared are stained with particular dyes and are resolved at the same time, on the same gel avoiding misinterpretation. Finally, 2D LC/MS-MS techniques arose combining directly the advantage of chemical derivatisation which label sample, and the separation of proteins unresolved by 2D gel, like very basic or acidic ones [16].

But in the context of clinical diagnosis, protein profiling on chip or after purification steps with chemically activated microbeads, MALDI imaging and related techniques are probably the more exciting methods. Protein profiling would allow the determination of particular signatures of proteins and peptides, to be informative for different pathological states. These instruments and techniques, like SELDI-TOF-MS or MALDI-TOF-MS profiling are therefore moving towards an adaptation to diagnosis in medical area. MALDI imaging will probably be a complement to classical histological analysis. It is based on protein

profiling directly on tissue section. Then, it may directly target total protein content of an embedded tissue portion with very little step of preparation and no need of presolubilization or fractionation of sample. Nevertheless, tissue preparation for MALDI analysis has to be optimized in order to enhance desorption of constituting molecules. Moreover, MALDI Imaging should also provide spatial information on proteins, with localised concentration as computer program can build 3 dimensions views of a piece of tissue, with its serial slices [28].

4. Perspectives and concluding remarques:

Proteomics is certainly a very broad field and the integration of all works achieved in the area will open new perspectives in the diagnosis of malignancies. The data generated should be shared and organised in order to establish "proteomes analysis of human", like it was realised for human genome sequencing. But, these information exchanges would need even more energy and organisation, as proteome is certainly broader then genome. Indeed, postranslational modifications, splicing isoforms of the same protein, specificity of tissue expressions, all together multiply and complicate input data. But the integration of genomic, transcriptomic and proteomic data with classical signalling activation information is probably indicated to meet the real complexity of a biological sample [29]. This kind of very ambitious work, increasing study in every fields and the development of new techniques will certainly implement our knowledge on CRC and will guide scientists towards new possibility for efficient clinical diagnosis and maybe screening of CRC.

1. Hakama, M., et al., *Screening for colorectal cancer.* Acta Oncol, 2005. **44**(5): p. 425-39.
2. Nicholson, F.B., et al., *Review article: Population screening for colorectal cancer.* Aliment Pharmacol Ther, 2005. **22**(11-12): p. 1069-77.
3. Ouyang, D.L., et al., *Noninvasive testing for colorectal cancer: a review.* Am J Gastroenterol, 2005. **100**(6): p. 1393-403.
4. Kahlenberg, M.S., et al., *Molecular prognostics in colorectal cancer.* Surg Oncol, 2003. **12**(3): p. 173-86.
5. Bendardaf, R., H. Lamlum, and S. Pyrhonen, *Prognostic and predictive molecular markers in colorectal carcinoma.* Anticancer Res, 2004. **24**(4): p. 2519-30.
6. Hemandas, A.K., et al., *Metastasis-associated protein S100A4--a potential prognostic marker for colorectal cancer.* J Surg Oncol, 2006. **93**(6): p. 498-503.
7. Kawakita, M. and K. Hiramatsu, *Diacetylated derivatives of spermine and spermidine as novel promising tumor markers.* J Biochem (Tokyo), 2006. **139**(3): p. 315-22.
8. Garcia, V., et al., *Thymidylate synthase messenger RNA expression in plasma from patients with colon cancer: prognostic potential.* Clin Cancer Res, 2006. **12**(7 Pt 1): p. 2095-100.
9. Wang, W.S., et al., *Plasma von Willebrand factor level as a prognostic indicator of patients with metastatic colorectal carcinoma.* World J Gastroenterol, 2005. **11**(14): p. 2166-70.
10. Ahlquist, D.A., J.E. Skoletsky, and K.A.Boynton, *Colorectal cancer screening by detection of alterd human DNA in stool: feasability of a multitarget assay panel.* Gatsroenterology, 2000. **119**: p. 1219-1227
11. von Kleist, S., Y. Hesse, and H. Kananeeh, *Comparative evaluation of four tumor markers, CA 242, CA 19/9, TPA and CEA in carcinomas of the colon.* Anticancer Res, 1996. **16**(4B): p. 2325-31.
12. Drake, R.R., et al., *Serum, salivary and tissue proteomics for discovery of biomarkers for head and neck cancers.* Expert Rev Mol Diagn, 2005. **5**(1): p. 93-100.
13. Lam, Y.W., et al., *Mass profiling-directed isolation and identification of a stage-specific serologic protein biomarker of advanced prostate cancer.* Proteomics, 2005. **5**(11): p. 2927-38.
14. Petricoin, E.F., D.K. Ornstein, and L.A. Liotta, *Clinical proteomics: Applications for prostate cancer biomarker discovery and detection.* Urol Oncol, 2004. **22**(4): p. 322-8.
15. Alaiya, A., M. Al-Mohanna, and S. Linder, *Clinical cancer proteomics: promises and pitfalls.* J Proteome Res, 2005. **4**(4): p. 1213-22.
16. Skandarajah, A.R., et al., *Proteomic analysis of colorectal cancer: discovering novel biomarkers.* Expert Rev Proteomics, 2005. **2**(5): p. 681-92.
17. Merchant, M. and S.R. Weinberger, *Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry.* Electrophoresis, 2000. **21**(6): p. 1164-77.
18. Petricoin, E.F. and L.A. Liotta, *SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer.* Curr Opin Biotechnol, 2004. **15**(1): p. 24-30.
19. Melle, C., et al., *Discovery and identification of alpha-defensins as low abundant, tumor-derived serum markers in colorectal cancer.* Gastroenterology, 2005. **129**(1): p. 66-73.
20. Semmes, O.J., et al., *Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility.* Clin Chem, 2005. **51**(1): p. 102-12.

21. Friedman, D.B., et al., *Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry.* Proteomics, 2004. **4**(3): p. 793-811.
22. Alfonso, P., et al., *Proteomic expression analysis of colorectal cancer by two-dimensional differential gel electrophoresis.* Proteomics, 2005. **5**(10): p. 2602-11.
23. Krieg, R.C., et al., *ProteinChip Array analysis of microdissected colorectal carcinoma and associated tumor stroma shows specific protein bands in the 3.4 to 3.6 kDa range.* Anticancer Res, 2004. **24**(3a): p. 1791-6.
24. Albrethsen, J., et al., *Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study.* BMC Cancer, 2005. **5**: p. 8.
25. Lane, C.S., et al., *Identification of cytochrome P450 enzymes in human colorectal metastases and the surrounding liver: a proteomic appraoch.* European journal of cancer, 2004. **40**: p. 2127-2134.
26. Melle, C., et al., *Detection and identification of heat shock protein 10 as a biomarker in colorectal cancer by protein profiling.* Proteomics, 2006. **6**(8): p. 2600-8.
27. Belluco, C., et al., *Kinase substrate protein microarray analysis of human colon cancer and hepatic metastasis.* Clin Chim Acta, 2005. **357**(2): p. 180-3.
28. Chaurand, P., et al., *Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry.* Proteomics, 2003. **3**(11): p. 2221-39.
29. Sagynaliev, E., et al., *Web-based data warehouse on gene expression in human colorectal cancer.* Proteomics, 2005. **5**(12): p. 3066-78.